The role of efficacy and effectiveness trials

A. John Rush
Duke-NUS Graduate Medical School Singapore, Jalan Bukit Merah, 169547 Singapore

Fleischhacker and Goodwin raise some important points about the roles of efficacy and effectiveness studies, in the context of reviewing recent large pragmatic trials. They initially suggest that, should the efficacy and effectiveness trials result in different conclusions, we have to trust the efficacy trials. They also note that findings from the efficacy trials, if positive, may be valid for a limited subset of patients (i.e., those who meet often restrictive entry criteria used in efficacy trials or those who are being treated under "research conditions" in which vigorous treatment is provided). Later in their conclusion, however, they are a bit more forgiving of effectiveness trials, saying they are as needed, and they even suggest requiring effectiveness trials before full market availability is allowed for the manufacturer. So which is it? Are effectiveness trials of use or not?

My view is in some but not complete agreement with theirs. I do not believe there are two types of trials (i.e., efficacy vs. effectiveness). Particular designs are formulated to answer specific questions. Different questions beget different designs. Different designs will give different answers as they should, since they are answering different questions.

The efficacy studies (Phase II - III) are designed with maximal internal validity to answer questions like: what is the efficacy, safety and tolerability of treatment "X" as compared to placebo (i.e., to isolate the clinical effects of the molecule or device *alone* on the patient (as expressed by side effects) and on the disease (as expressed as therapeutic or worsening effects).

Effectiveness trials entail a host of different designs, that address a range of different questions. Specifically, as noted by Fleischhacker and Goodwin, these trials enroll a wider range of patients, employ a wider range of "clinically relevant" outcomes, and provide treatment under "usual" vs. "research" conditions (which may or may not increase retention whilst risking underdosing). When an efficacy trial reveals efficacy, the magnitude of the effect may well be different in practice, depending on *who* is being treated (i.e., which patients) and *how* they are being treated. For depression, patients with anxious symptom features may do less well (1) than less anxious patients, even when equivalently treated. Indeed, as STAR*D showed, the time at which, in a sequence of interventions, a treatment is used will affect the chances of remission (2).

I am in substantial agreement with their conclusion: that both "types" of trials are useful. Each contributes to our understanding. No one design provides a unique path to the truth. Rather, the first question is: does the potential treatment actually work for not very complicated patients and at what cost or risk (e.g. adverse events) to patients? If the benefit outweighs the negative effects (as evaluated in efficacy trials), then where, how, and for whom is the treatment to be recommended? These latter questions are partially addressed by so called "effectiveness" trials. The designs may hold constant or allow variance in the nature and types of patients (e.g., comorbidities, concomitant medications, etc.), treatment procedures (e.g., visit frequency, dose to titration, etc.), where the treatment is used in a sequence of treatments (i.e., levels of treatment resistance), etc.

In addition, effectiveness trials can address other practical issues. In STAR*D, for example, patients could select among treatment strategies. Substantial numbers chose to augment, while others chose to switch treatments. The different switch and augmentation medication treatments did not differ in remission rates – a clear answer to the question about the comparative efficacy of "in class", "out of class" or "dual action" agents as second step switch treatments, for example. That far fewer patients chose to both switch and augment is not surprising. Patients with high side effects and poor efficacy from step 1 would logically want a switch. Those with some benefits and tolerable side effects from step 1 would logically not want to lose the benefit and therefore preferred augmentation. What was most interesting was the STAR*D finding that greater levels of resistance have a major effect on outcomes – both

acutely and in follow up. Thus, when a treatment is used is as important as how and for which patients it is to be used. This finding should affect the subsequent designs in efficacy trials.

Efficacy studies can only evaluate efficacy under specific conditions. If effectiveness studies differ in outcomes, then logically, it is *not* the case that the treatment will *never* work. Rather, the treatment is likely to work only under select conditions defined by patient subgroups, treatment methods, or at where in the sequence of treatments the treatment is used for example.

In conclusion, each type of trial (efficacy/effectiveness) provides essential contributions to how best to treat our patients.

References

1. Fava M, Rush AJ, Alpert JE et al. Difference in treatment outcome in outpatients with anxious versus nonanxious depression: a STAR*D report. Am J Psychiatry 2008; 165:342-51.
2. Rush AJ, Trivedi MH, Wisnieski SR et al. Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR*D report. Am J Psychiatry 2006;163:1905-17.